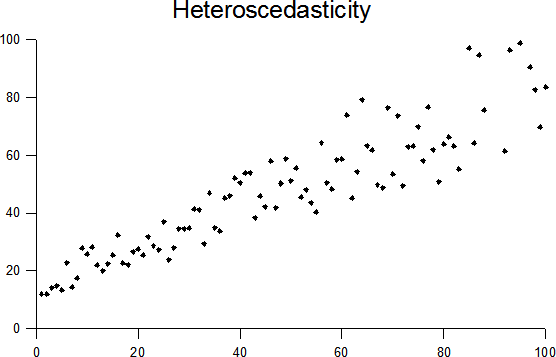
**Multiple Choice/Fill in the blank/Open-ended questions**

Circle only one choice or fill in only one word for each question unless otherwise noted as an open-ended question.

1. Say I develop a linear model of the functional form y ~ x1 + x2 + x3. Can I use the linear model to do the following (4 points):
   1. Predict variable y based on variables x1, x2, and x3. TRUE/FALSE
   2. Examine the effect of a change in variable x1 on a change in y independent of the effect of variables x2 and x3. TRUE/FALSE
   3. Examine which variables x1, x2, and x3 are significantly associated with y after controlling for others. TRUE/FALSE
   4. Identify the relative importance of x1, x2, and x3 in explaining variation in y when standardizing each x variable by its standard deviation and mean. TRUE/FALSE
2. Please draw your best estimate of the regression line through the scatter of points below (1 point):



1. Does it seem appropriate to use a linear model to fit the data shown in Question 2? Why or why not (1 point; open-ended)? No, errors are heteroscedastic
2. Say that I develop a linear model that estimates bird diversity per m2 based on the diversity of trees in a forest per m2, the ratio of edge to total forest cover, and the mean amount of rainfall the forest receives per year using data from North American forests in the year 2015.

Bird diversity (per m2) ~ diversity of trees in a forest (per m2) + ratio of forest edge to total forest cover + mean rainfall (mm per year)

I now want to use this model to predict the amount of bird diversity for forest patches in South America where I have data on the diversity of trees per m2, the ratio of edge to total forest cover, and the mean amount of rainfall in the year 2016. What assumptions do I make when using the linear model I developed for North America to predict bird diversity in this new system (3 points)?

* 1. I assume that the relationship between my 3 predictor variables and bird diversity is the same in North and South America. TRUE/FALSE.
  2. I assume that the sample size of data I have for North and South America are the same. TRUE/FALSE.
  3. I assume that the total amount of rainfall experienced in 2016 is exactly the same as the total amount of rainfall experienced in 2015. TRUE/FALSE.

1. Do you think it is appropriate to use the linear model developed to predict bird diversity in North America to predict bird diversity in South America (the example described in question 4)? Why or why not (2 points)?

I lean towards saying no because there are probably lots of differences between these systems that cant be accounted for in the model but I think we should accept either answer as long as the students reasoning is sound.

I agree.

1. Which of the following are true about a linear model and its assumptions. If helpful, you can assume the model takes the form of y ~ x (8 points)?
   1. There must be a linear relationship between your independent variable (x) and your dependent variable (y) TRUE/FALSE.
   2. The errors are heteroscedastic TRUE/FALSE
   3. There must be a significant relationship between your independent variable (x) and your dependent variable (y) TRUE/FALSE.
   4. Your errors do not have to be normally distributed as long as your sample size is large enough TRUE/FALSE.
   5. There is constant variance of the errors from your model TRUE/FALSE
   6. The errors must be normally distributed TRUE/FALSE
   7. The errors must be statistically independent TRUE/FALSE
   8. If you violate the assumptions of a linear model it may result in inaccurate estimates of confidence intervals but your p values should be correct TRUE/FALSE
2. Which of the following is true about statistical power (4 points)?
   1. It is a measure of Type I error. TRUE/FALSE
   2. One way to increase statistical power is by allowing for a smaller p value to determine significance (e.g. p < 0.01 instead of p < 0.05). TRUE/FALSE
   3. It is independent of the effect size of your variable of interest relative to the variability of the response variable (?) TRUE/FALSE
   4. One way to increase statistical power is by increasing your sample size TRUE/FALSE
3. Which of the following statements are true considering model selection (5 points)?
   1. You can use an F test to compare non-nested models. TRUE/FALSE
   2. If you select the best candidate modeling using AIC, you should select the model with the highest AIC score. TRUE/FALSE.
   3. You should select a few candidate models based on which variable combinations you think are important in theory and compare these models using model selection techniques TRUE/FALSE
   4. You should always use model selection in favor of using a full model. TRUE/FALSE.
   5. You do not have to worry about multicollinearity when doing model selection because model selection techniques always pick the model with the least multicollinearity. TRUE/FALSE
4. Which of the following statements are true considering multicollinearity of predictor variables x1 and x2 in a linear model of the functional form (4 points):

y ~ x1 + x2

* 1. When you put two highly correlated variables, x1 and x2, in a linear model, the standard errors and p values for x1 and x2 are larger than they would be for x1 and x2 in two separate univariate (single variable) models TRUE/FALSE
  2. When you put two highly correlated variables, x1 and x2, in a linear model, the beta coefficients for x1 and x2 will likely be the same when compared to beta coefficients produced in two separate univariate (single variable) models TRUE/FALSE
  3. You will be unable to accurately predict y using your linear model if you include both x1 and x2 TRUE/FALSE
  4. It will be difficult to identify the true effect of x1 and x2 on y. TRUE/FALSE

**Open-ended Questions:**

1. Say that you are interested in designing a study to examine the impact of increased pollination on crop yields. How would you design an experimental study to test this? How would you design an observational study? Please address the following in each of your answers: (1) sampling strategy, (2) sample size and power, and (3) ways to reduce the effect of confounding factors (6 points total).
   1. Experimental (3 points)
   2. Observational (3 points)
2. 2) Please interpret the linear regression model output below

* How does the interpretation of the beta coefficients change when you add an interaction term
* interpret the following regression output after log transforming your x variable.
* Interpret the output of this regression result that includes both continuous and categorical predictor variables (e.g. ANCOVA).